FIG. 1

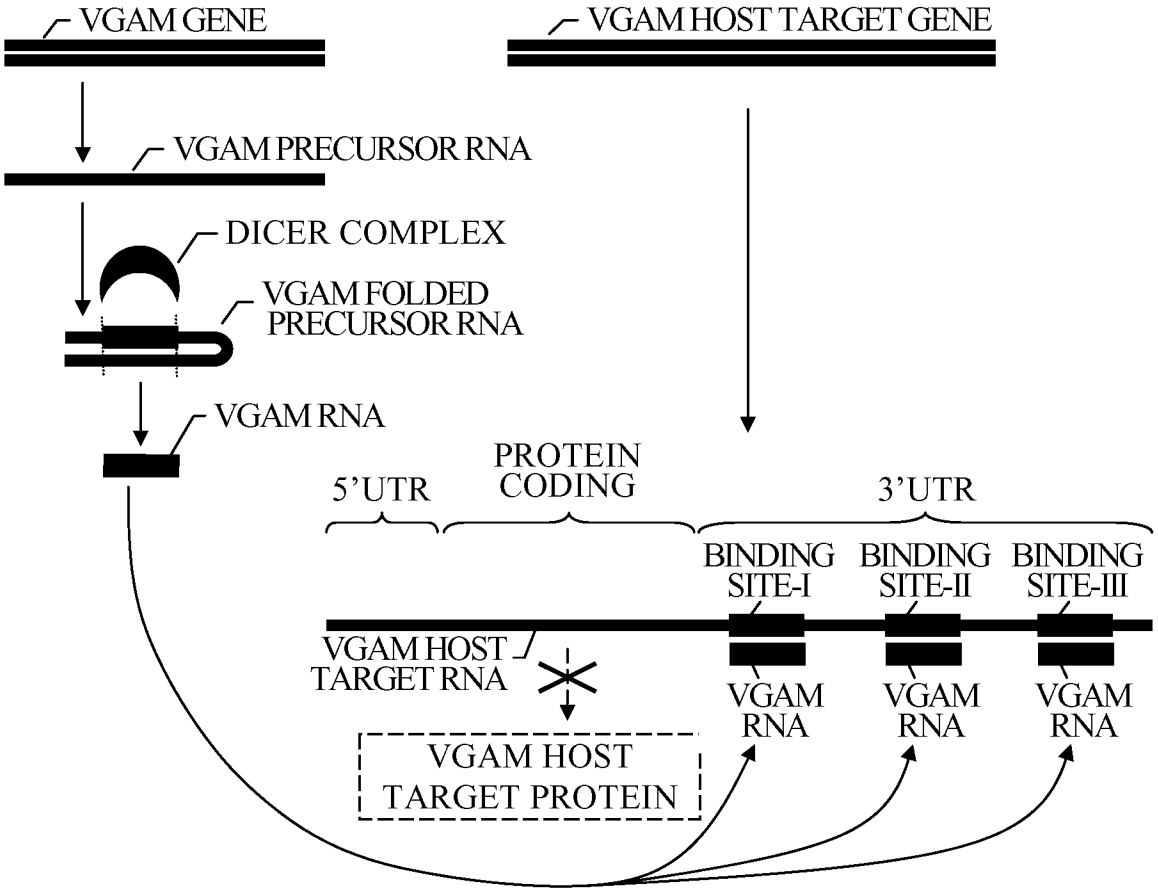

VGAM GENE
VGAM HOST TARGET GENE
VGAM PRECURSOR RNA
DICER COMPLEX
VGAM FOLDED PRECURSOR RNA
VGAM RNA
PROTEIN CODING
5'UTR
3'UTR
BINDING SITE-I
BINDING SITE-II
BINDING SITE-III
VGAM HOST TARGET RNA
VGAM RNA
VGAM RNA
VGAM RNA
VGAM HOST TARGET PROTEIN

FIG. 2

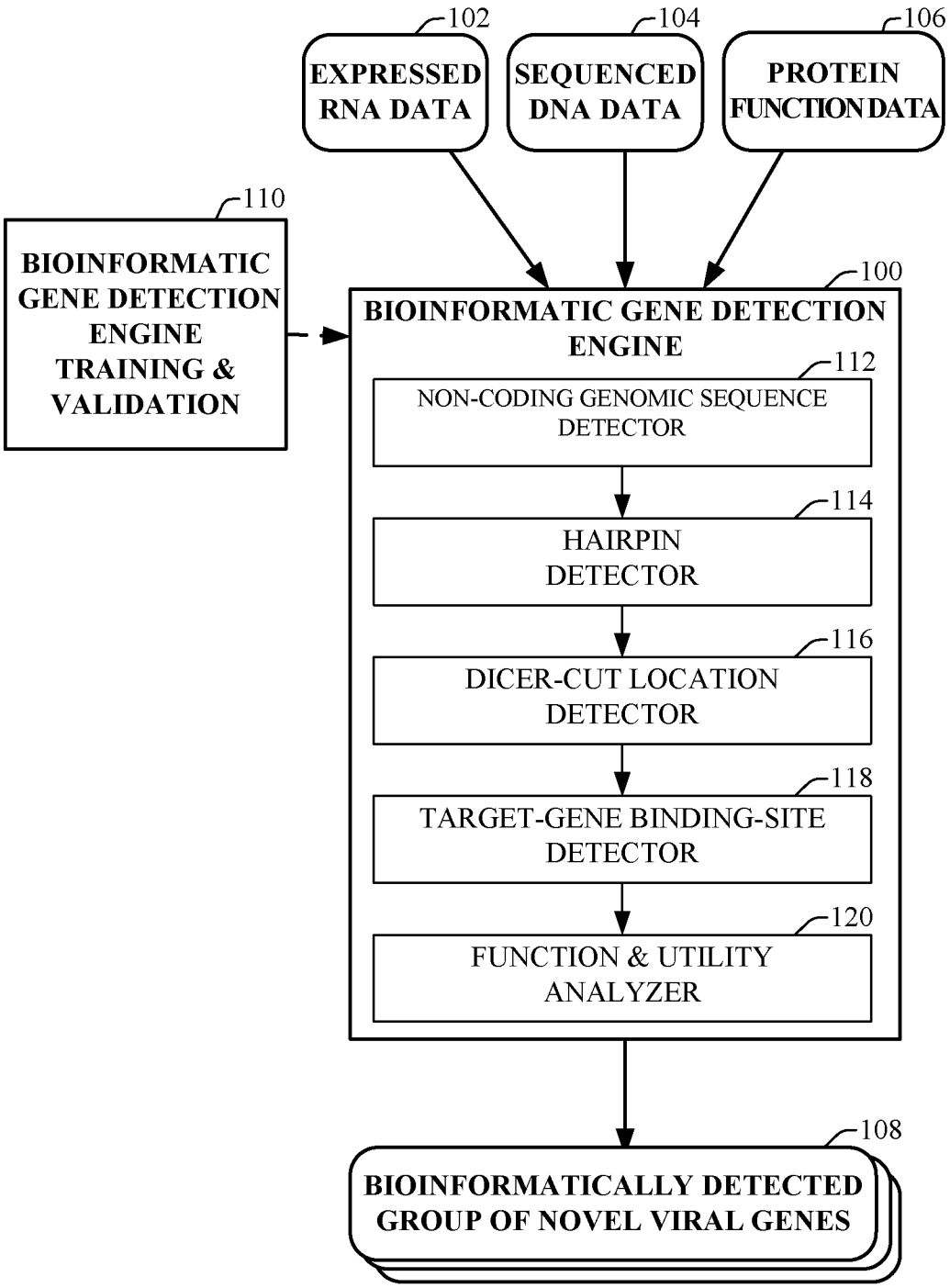

102
104
106
EXPRESSED RNA DATA
SEQUENCED DNA DATA
PROTEIN FUNCTION DATA
110
BIOINFORMATIC GENE DETECTION ENGINE TRAINING & VALIDATION
100
BIOINFORMATIC GENE DETECTION ENGINE
112
NON-CODING GENOMIC SEQUENCE DETECTOR
114
HAIRPIN DETECTOR
116
DICER-CUT LOCATION DETECTOR
118
TARGET-GENE BINDING-SITE DETECTOR
120
FUNCTION & UTILITY ANALYZER
108
BIOINFORMATICALLY DETECTED GROUP OF NOVEL VIRAL GENES

FIG. 3

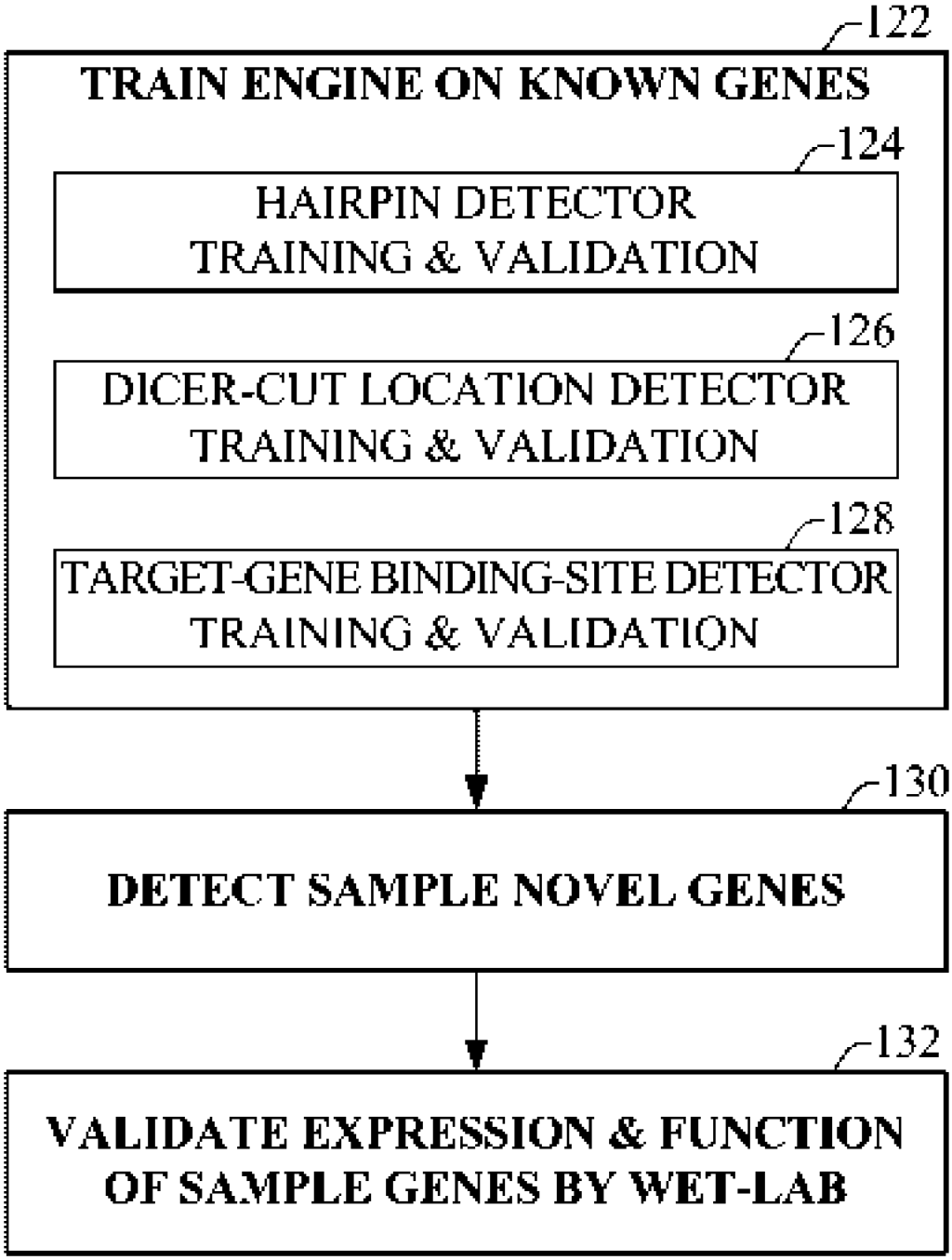

122
TRAIN ENGINE ON KNOWN GENES
124
HAIRPIN DETECTOR TRAINING & VALIDATION
126
DICER-CUT LOCATION DETECTOR TRAINING & VALIDATION
128
TARGET-GENE BINDING-SITE DETECTOR TRAINING & VALIDATION
130
DETECT SAMPLE NOVEL GENES
132
VALIDATE EXPRESSION & FUNCTION OF SAMPLE GENES BY WET-LAB

FIG. 4A

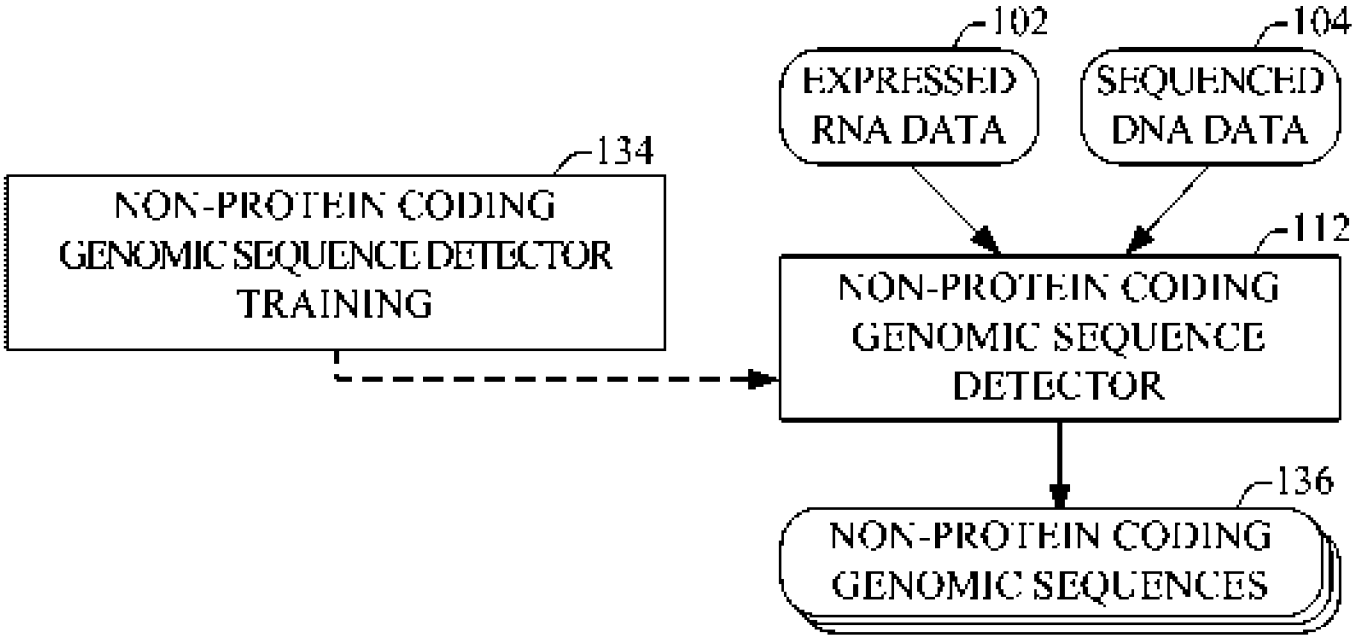

102
EXPRESSED RNA DATA
104
SEQUENCED DNA DATA
134
NON-PROTEIN CODING GENOMIC SEQUENCE DETECTOR TRAINING
112
NON-PROTEIN CODING GENOMIC SEQUENCE DETECTOR
136
NON-PROTEIN CODING GENOMIC SEQUENCES

FIG. 4B

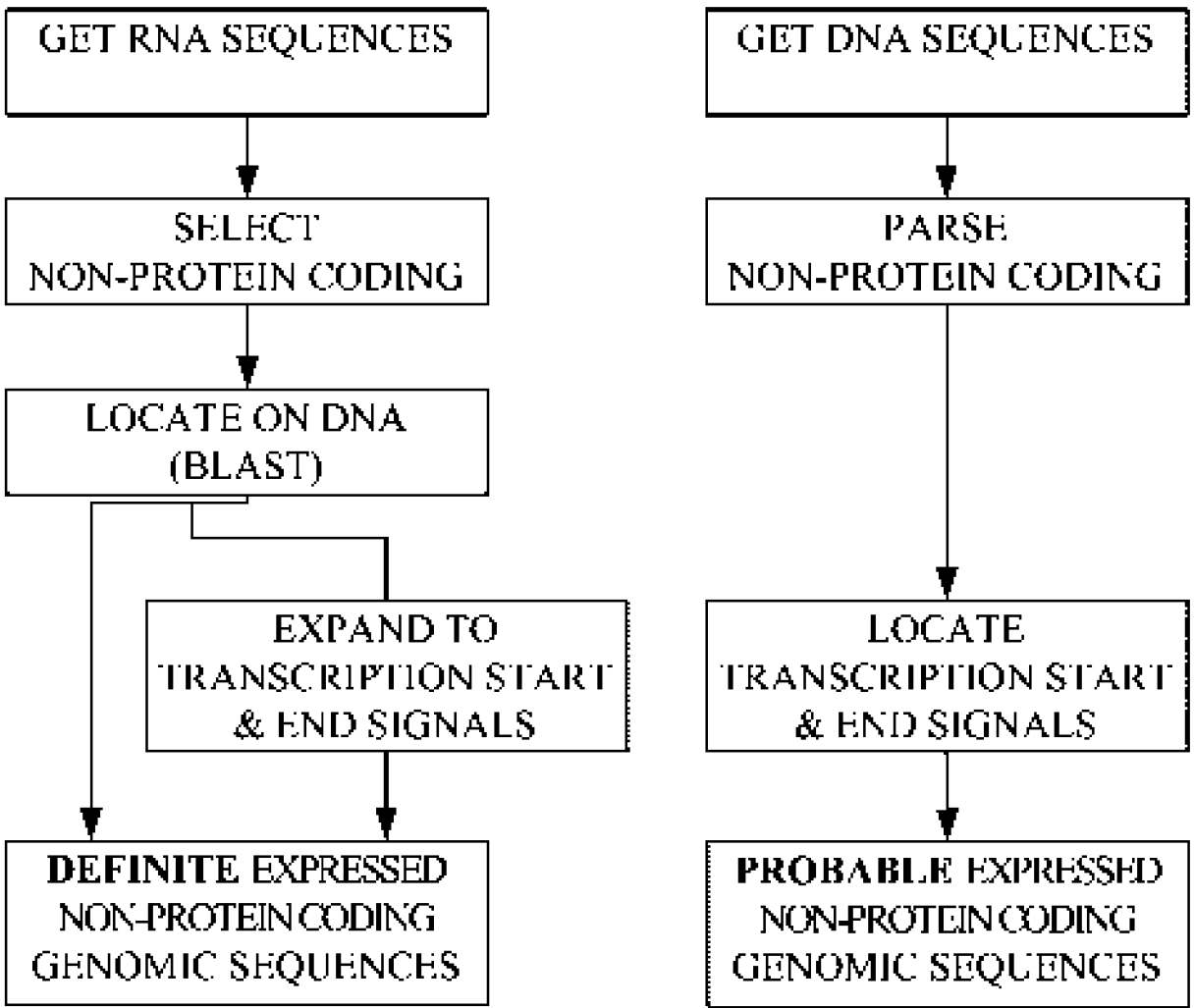

GET RNA SEQUENCES
SELECT NON-PROTEIN CODING
LOCATE ON DNA (BLAST)
EXPAND TO TRANSCRIPTION START & END SIGNALS
DEFINITE EXPRESSED NON-PROTEIN CODING GENOMIC SEQUENCES
GET DNA SEQUENCES
PARSE NON-PROTEIN CODING
LOCATE TRANSCRIPTION START & END SIGNALS
PROBABLE EXPRESSED NON-PROTEIN CODING GENOMIC SEQUENCES

FIG. 5A

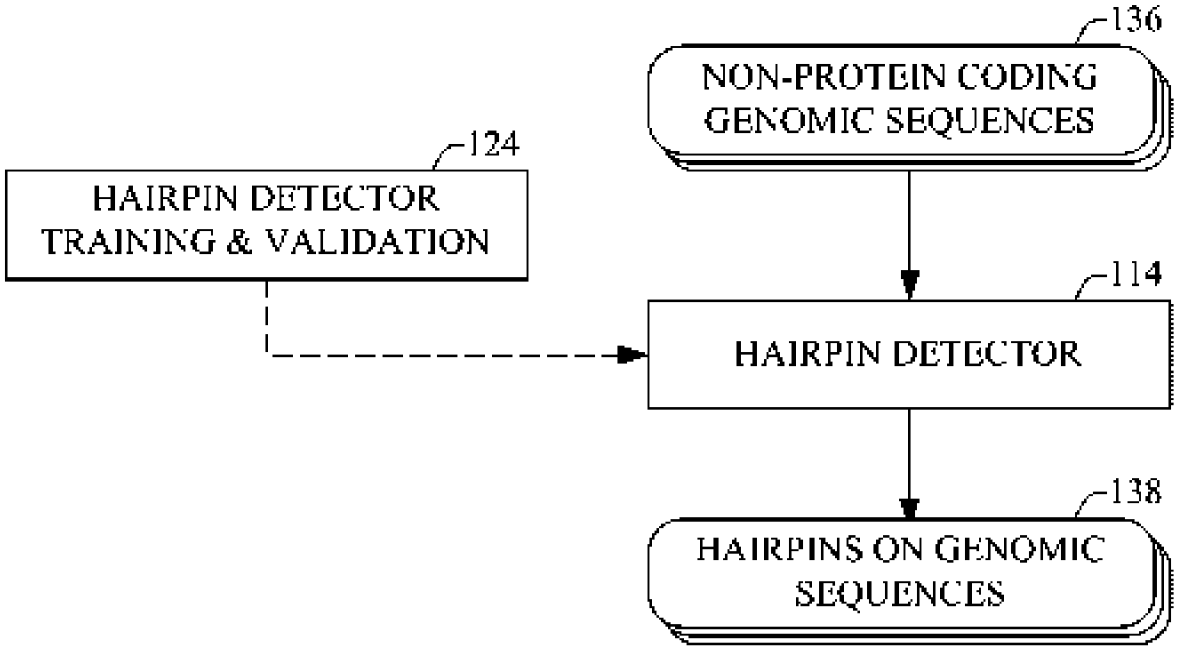

136
NON-PROTEIN CODING GENOMIC SEQUENCES
124
HAIRPIN DETECTOR TRAINING & VALIDATION
114
HAIRPIN DETECTOR
138
HAIRPINS ON GENOMIC SEQUENCES

FIG. 5B

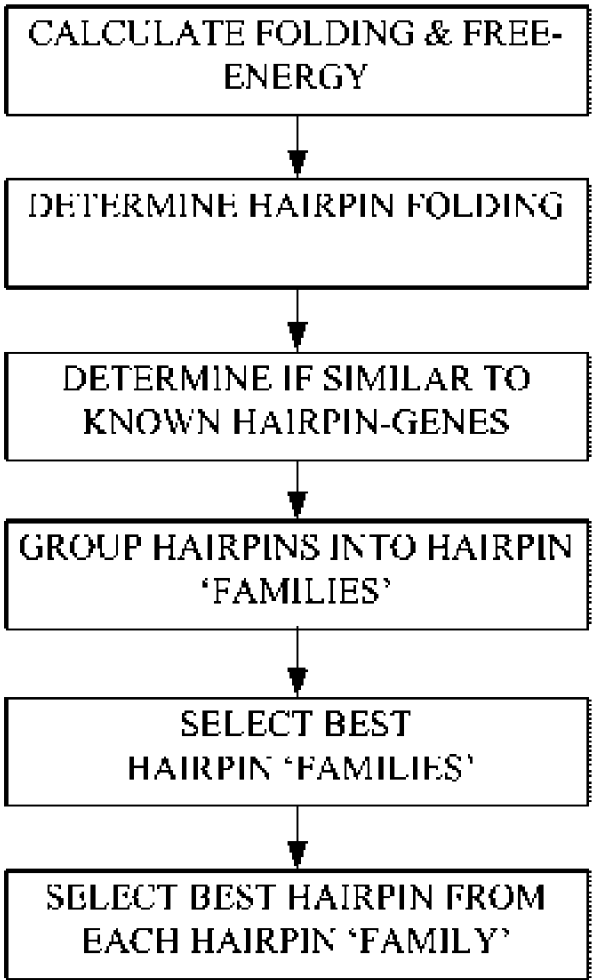

CALCULATE FOLDING & FREE-ENERGY
DETERMINE HAIRPIN FOLDING
DETERMINE IF SIMILAR TO KNOWN HAIRPIN-GENES
GROUP HAIRPINS INTO HAIRPIN 'FAMILIES'
SELECT BEST HAIRPIN 'FAMILIES'
SELECT BEST HAIRPIN FROM EACH HAIRPIN 'FAMILY'

FIG. 6A

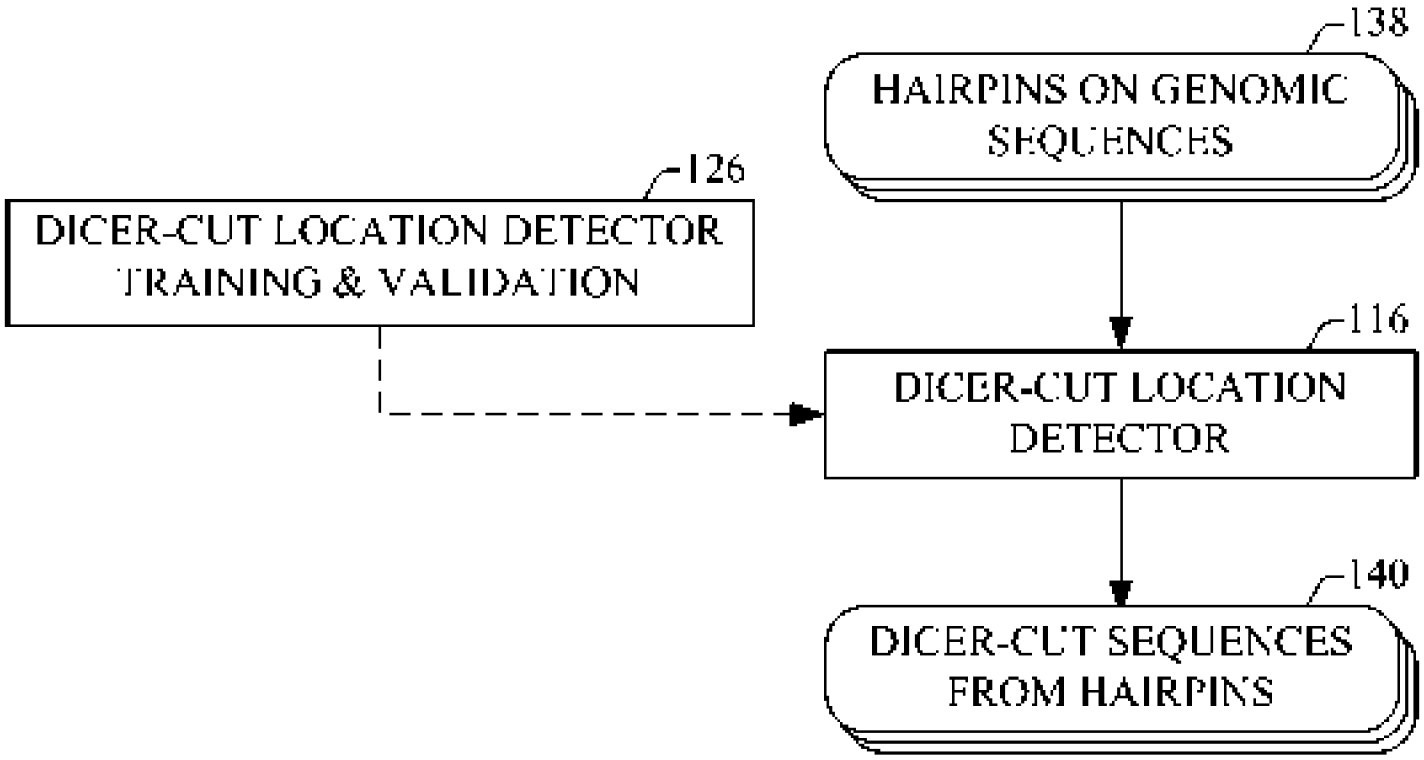

138
HAIRPINS ON GENOMIC SEQUENCES
126
DICER-CUT LOCATION DETECTOR TRAINING & VALIDATION
116
DICER-CUT LOCATION DETECTOR
140
DICER-CUT SEQUENCES FROM HAIRPINS

FIG. 6B

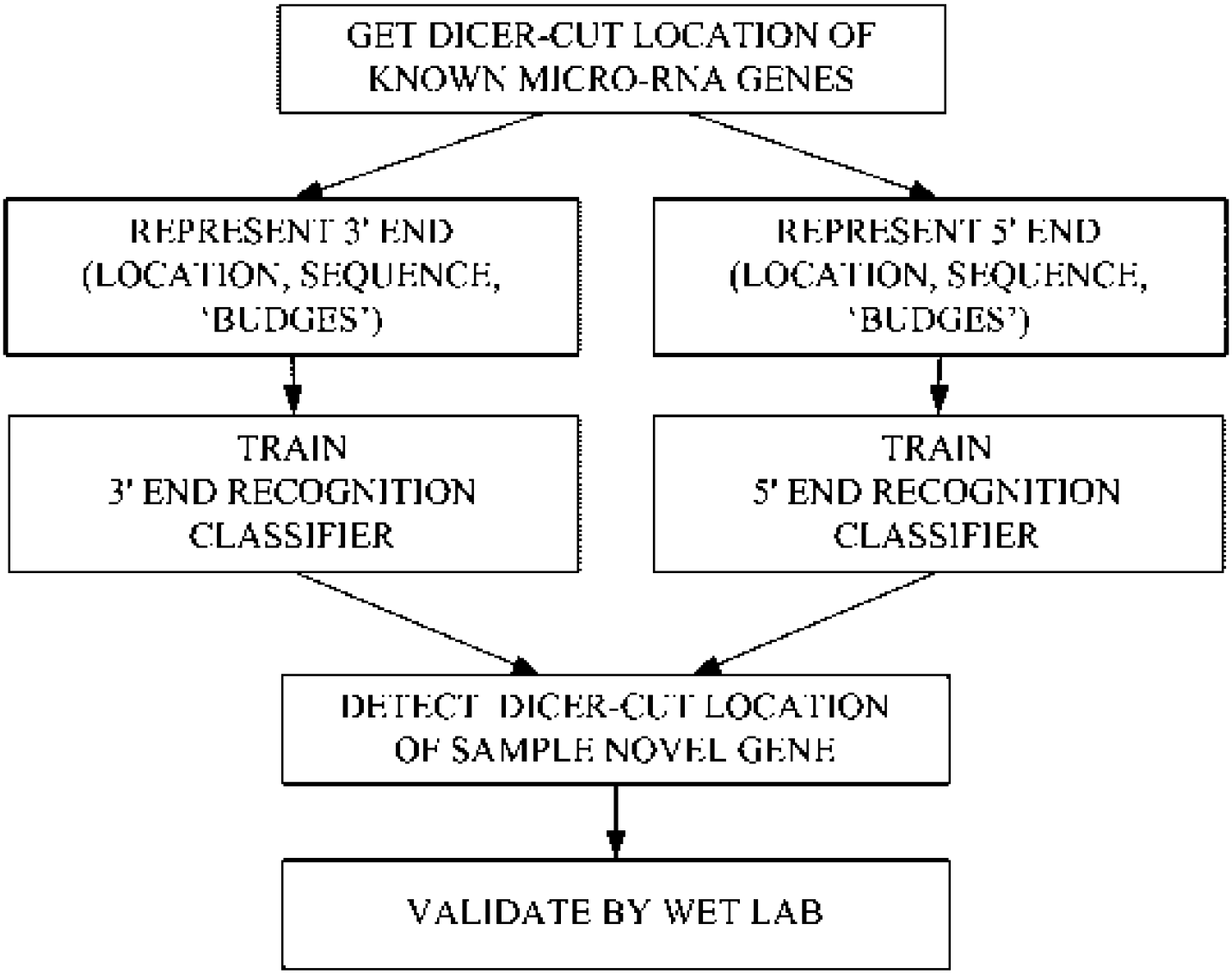

GET DICER-CUT LOCATION OF KNOWN MICRO-RNA GENES
REPRESENT 3' END (LOCATION, SEQUENCE, 'BUDGES')
REPRESENT 5' END (LOCATION, SEQUENCE, 'BUDGES')
TRAIN 3' END RECOGNITION CLASSIFIER
TRAIN 5' END RECOGNITION CLASSIFIER
DETECT DICER-CUT LOCATION OF SAMPLE NOVEL GENE
VALIDATE BY WET LAB

FIG. 6C

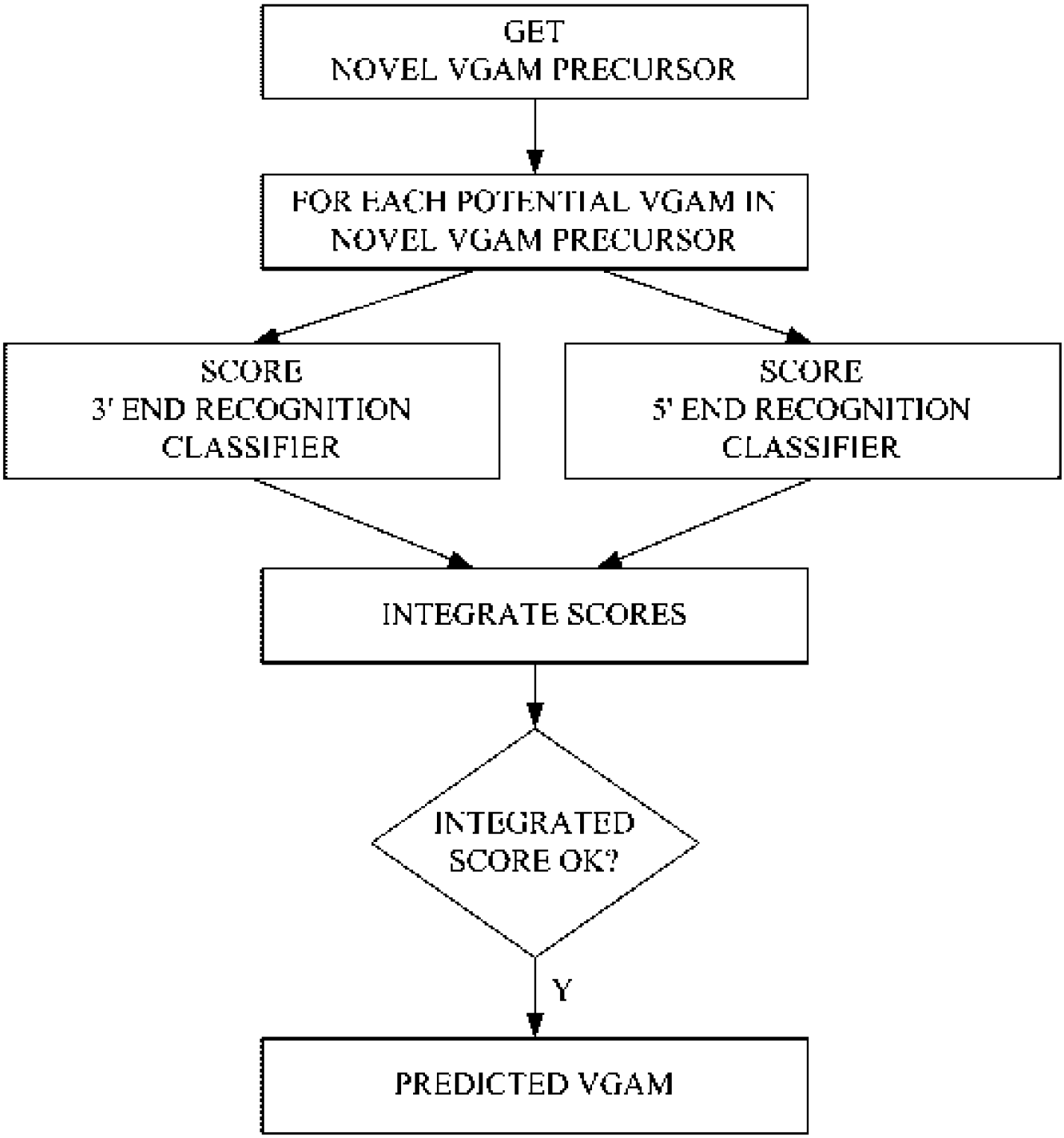

GET
NOVEL VGAM PRECURSOR
FOR EACH POTENTIAL VGAM IN
NOVEL VGAM PRECURSOR
SCORE
3' END RECOGNITION
CLASSIFIER
SCORE
5' END RECOGNITION
CLASSIFIER
INTEGRATE SCORES
INTEGRATED
SCORE OK?
Y
PREDICTED VGAM

FIG. 7A

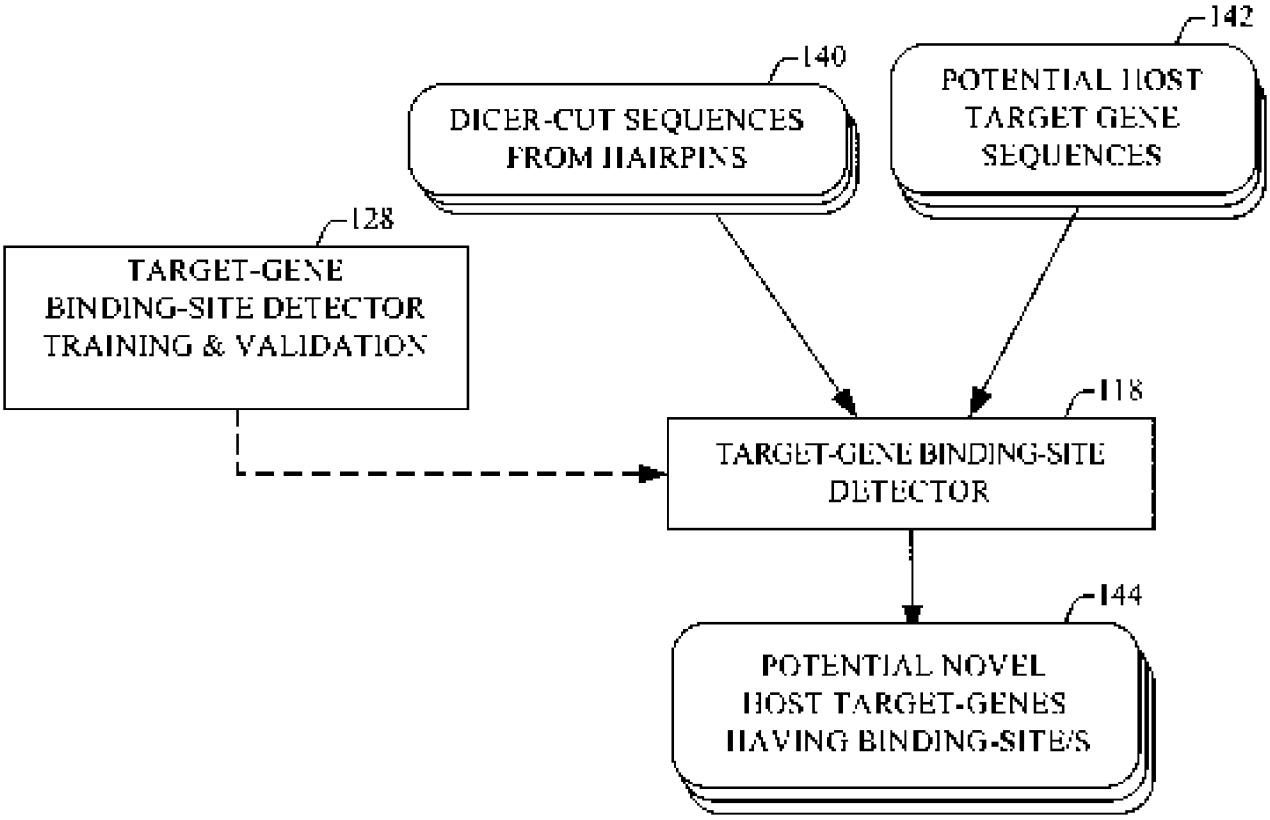

142
140
DICER-CUT SEQUENCES FROM HAIRPINS
POTENTIAL HOST TARGET GENE SEQUENCES
128
TARGET-GENE BINDING-SITE DETECTOR TRAINING & VALIDATION
118
TARGET-GENE BINDING-SITE DETECTOR
144
POTENTIAL NOVEL HOST TARGET-GENES HAVING BINDING-SITE/S

FIG. 7B

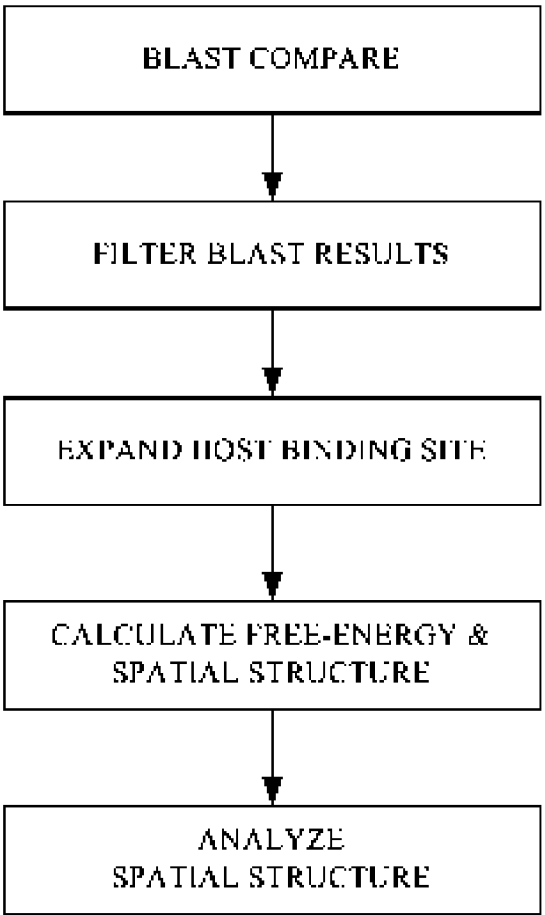

BLAST COMPARE
FILTER BLAST RESULTS
EXPAND HOST BINDING SITE
CALCULATE FREE-ENERGY & SPATIAL STRUCTURE
ANALYZE SPATIAL STRUCTURE

FIG. 8

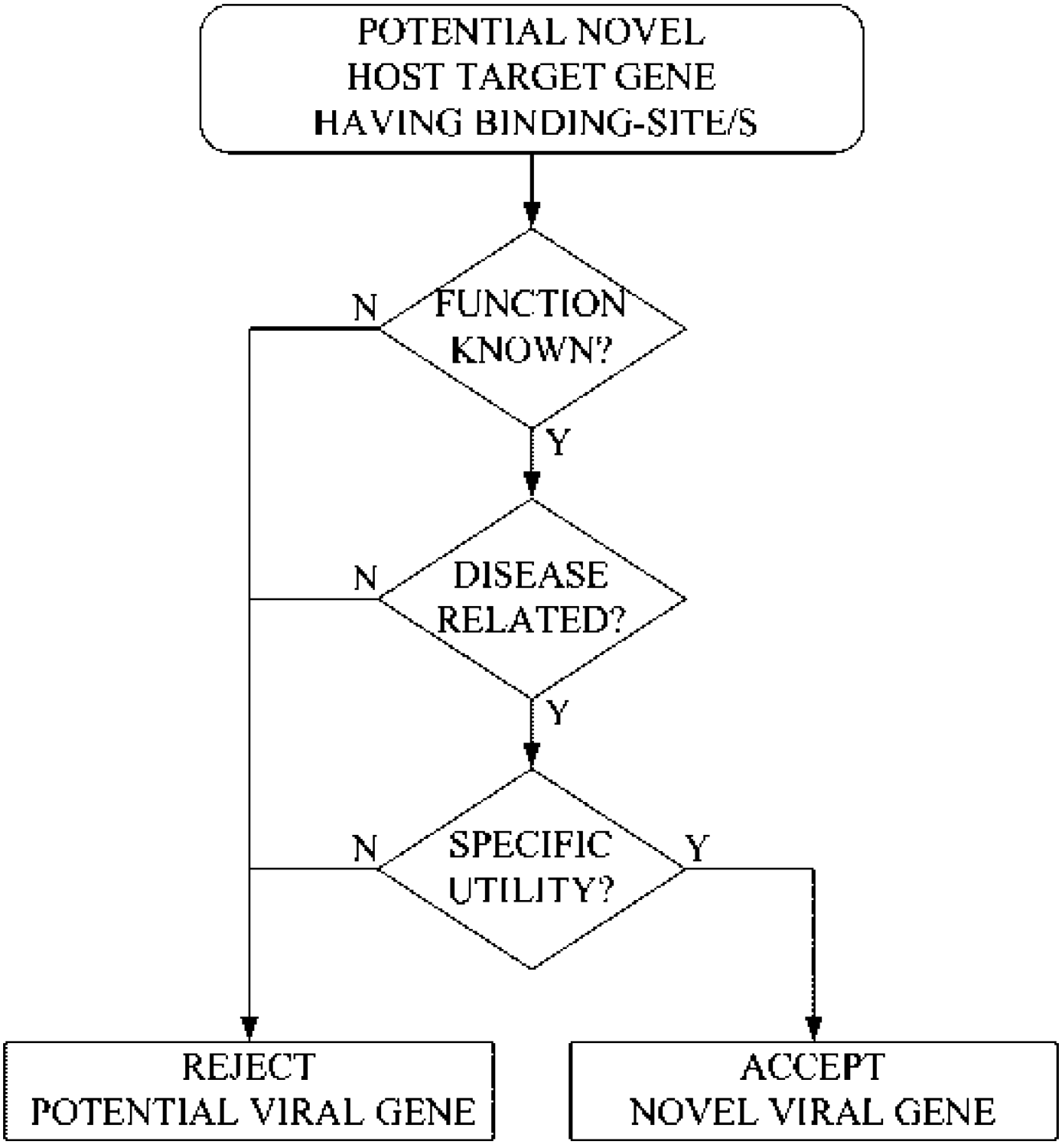

POTENTIAL NOVEL
HOST TARGET GENE
HAVING BINDING-SITE/S
FUNCTION
KNOWN?
N
Y
DISEASE
RELATED?
N
Y
SPECIFIC
UTILITY?
N
Y
REJECT
POTENTIAL VIRAL GENE
ACCEPT
NOVEL VIRAL GENE

FIG. 9

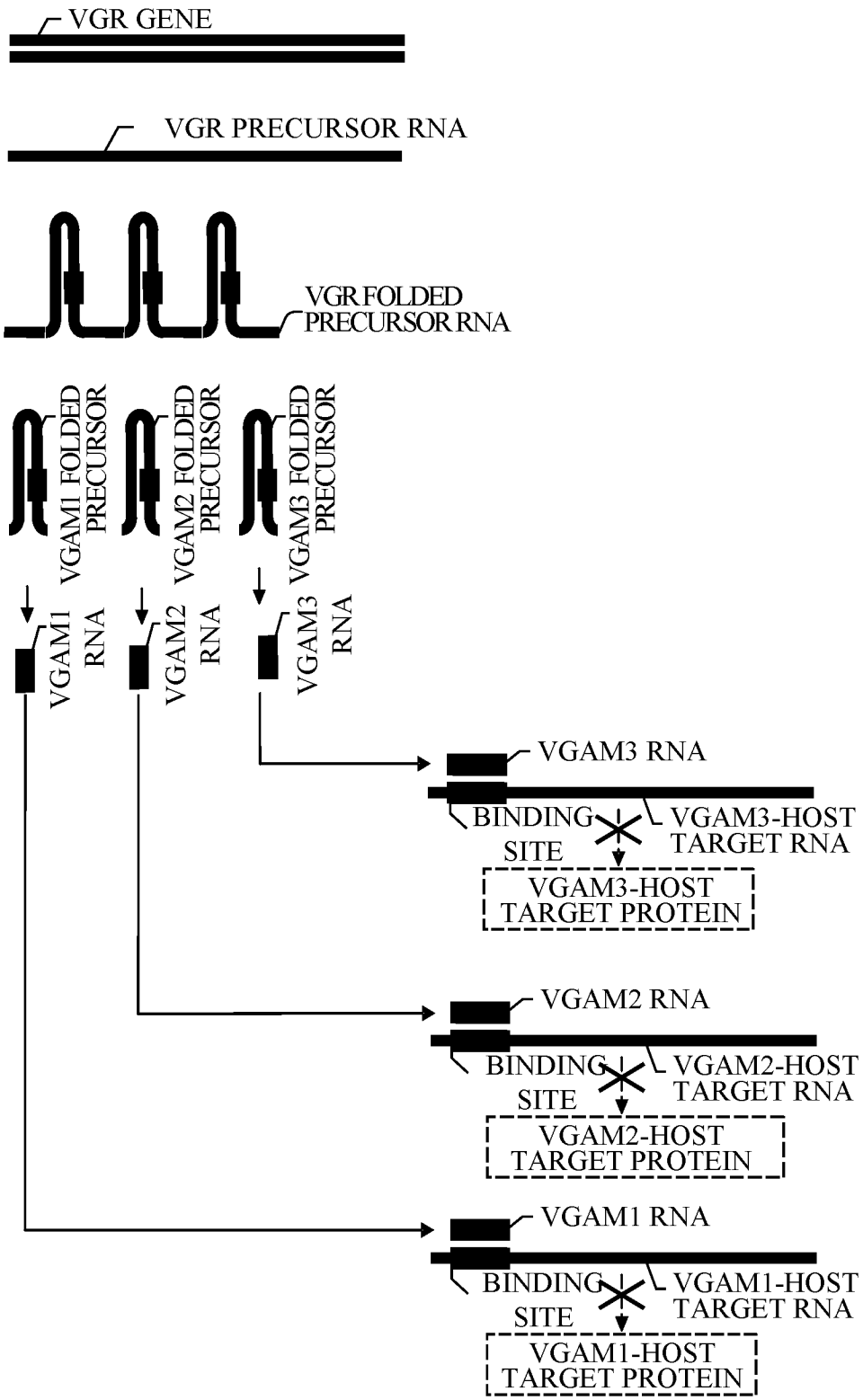

VGR GENE
VGR PRECURSOR RNA
VGR FOLDED PRECURSOR RNA
VGAM1 FOLDED PRECURSOR
VGAM2 FOLDED PRECURSOR
VGAM3 FOLDED PRECURSOR
VGAM1 RNA
VGAM2 RNA
VGAM3 RNA
VGAM3 RNA
BINDING SITE
VGAM3-HOST TARGET RNA
VGAM3-HOST TARGET PROTEIN
VGAM2 RNA
BINDING SITE
VGAM2-HOST TARGET RNA
VGAM2-HOST TARGET PROTEIN
VGAM1 RNA
BINDING SITE
VGAM1-HOST TARGET RNA
VGAM1-HOST TARGET PROTEIN

FIG. 10

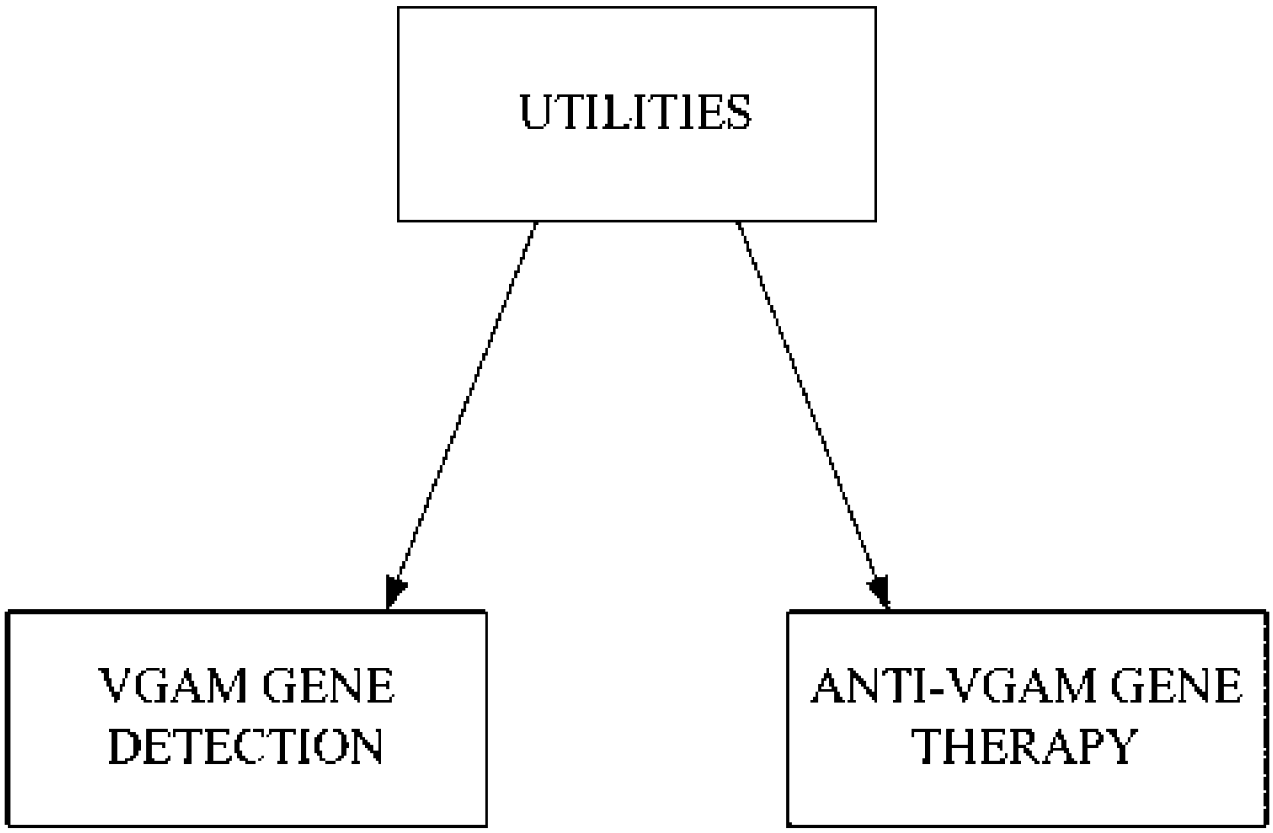


FIG. 11A

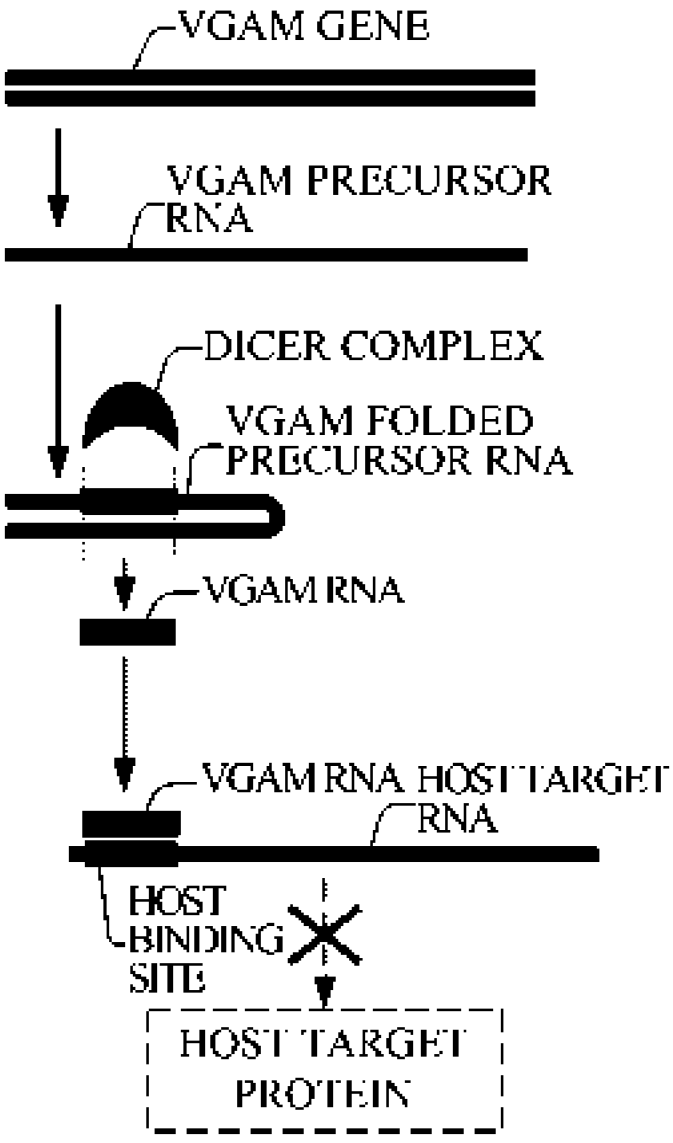
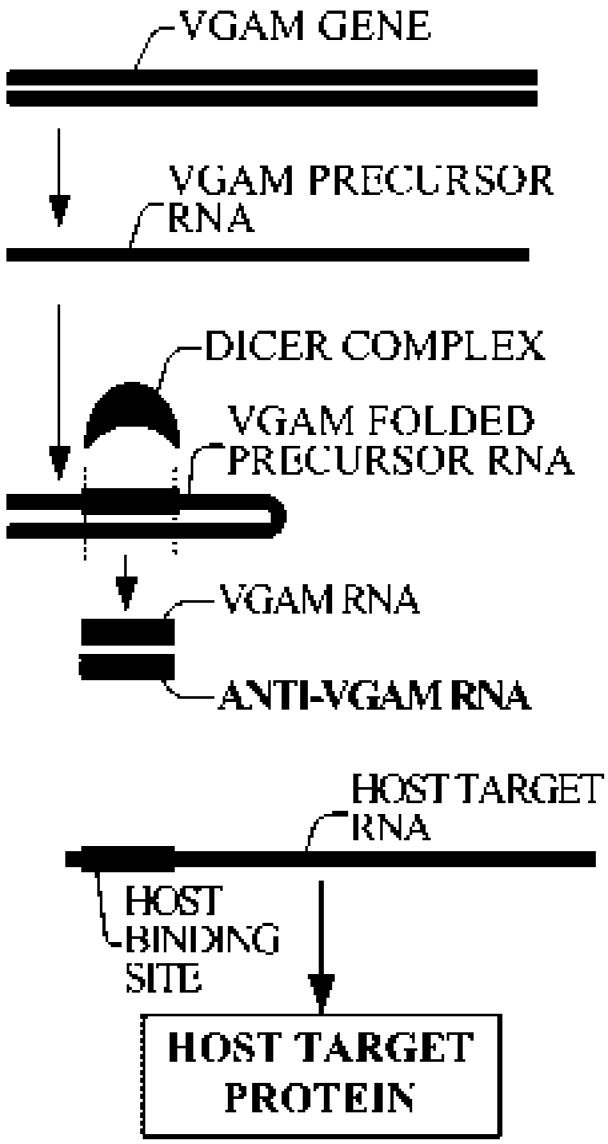


FIG. 11B



EST72223 sequence:

FIG. 12A

CCCTTATTAGAGGATTCTGCTCATGCCAGG**GTGAGGTAGTAAGTTGT**
ATTGTTGTGGGGTAGGGATATTAGGCCCCAATTAGAAGATAACTAT
ACAACTTACTACTTTCCCTGGTGTGTGGCATATTCACACTTAGTCTTA MIR98
GCAGTGTTGCCTCCATCAGACAAAGTTGTAGATGTTCCTTGGATAATT
TGGACTGGAAGAAAAGAGACATGGAAGGGGACAGATGGTGTTTAGG
GTGAGGCAGATGTCATTATAAAGTGACTTGTCTTTCATTAATTGGAGC
ATATAATTATTTTACCTTTGGGCATGAACTCATTTTGCTATTCTTCAAC
TGTGTAATGATTGCATTTTATTAGTAATAGAACAGGAATGTGTGCAAG
GGAATGGAAAGCATACTTTAAGAATTTTGGGCCAGGCGCGGTGGTTC
ATGCCTGTAATCCCAGCATTTTTGGGAGGCCGAGGCGGGTGGATCA
CCTGAGGTCAGGAGTTCGAGACCAACCTGGCCAACACGGCGAAACC
CCGCCTCTACTCAAATACAAAAATTAGCCAGGCTTGGTGACACTCGC
CTGTGGTCCCAGCTACTCAGGAGGCT**GAGGCAGGAGAATTGCTTGA**
ACCCAGGAAGTGGAGGCTTCAGTGAGCTGAGAACACGCCACTGCA GAM24
CTCCAGTCCTGGGCAACAGAGCAAGACTCTGTCTCAGGAAAAAAAA
AG

FIG. 12B

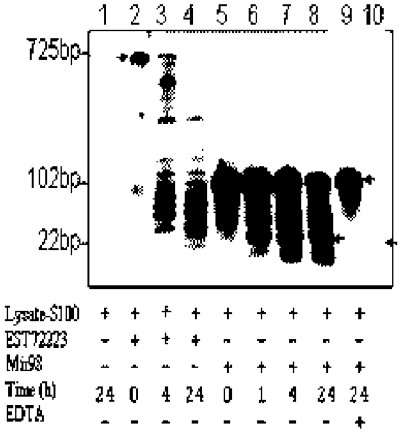
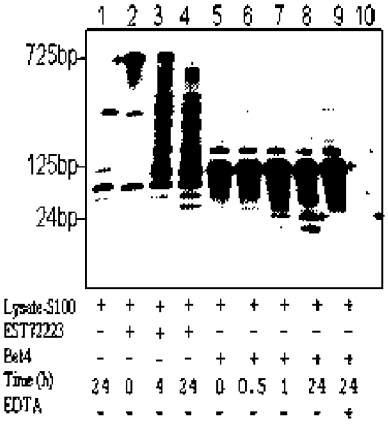


FIG. 12C



MIR98 GAM24

FIG. 12D

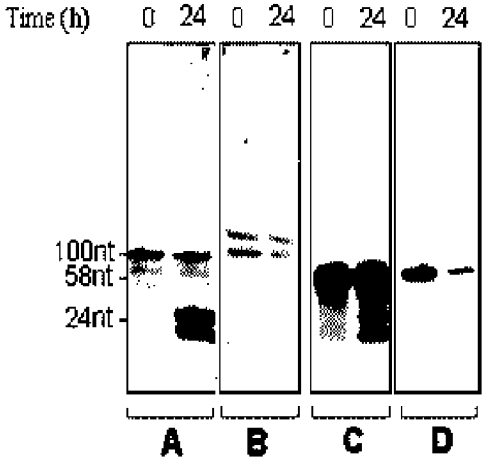


FIG. 13A

dbEST Id.7929020(Image4514344) sequence:

GCAAAAACTGGAAGCATTCCCTTTGAAAACTGGCACAAGACAGGGATGCCCTCT
CTCACCCCTCCTATTCAACATAGTCTTGGAAGTTCTGCCCAGGGCAATTAGGCA
GGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGAGCAAGTCAAATTGTTCCT
GTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGTCTCAGCCCCAAA
TCTCCTTAACCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAATAAATGT
ACAAAAATCACAAGCATTCTTACACACCAACAACAGAAAAACAGAGCCAAATCA
TGAGTGAACTCCCATTCACAATTGCTTCAAAGAGAATAAAATACCTAGGAATCC
AACTTACAAGGGATGTGAAGGACCTCTTCAAGGAGAACTACAAACCACTGCTCA
AGGAAATAAAAGAGGATACAAACAAATGGAAGAACATTCCATGCTCATGGGTAG
GAAGAATCAATATTGTGAAAATGGCCATACTGCCCAAGGTAATTTACAGATTCA
ATGCCATCCCCATCAAGCTACCAATGACTTTCTTCACAGAATTGGAAAAAACTA
CTTTAAAGTTCATATGGAACCAAAAAAGAGCCCGCATCGCCAAGTCAATCCTAA
GCCAAAAGAACAAAGCTGGAGGCATCACACTACCTGACTTCAAACTTTACTACA GAM23
AGGCTACAGTAACCAAAACAGCATGGTACTGGTACCAAAACAGACATATAGATC
AATGGAACAGAACAGAGCCCTCAGAAATAACGCCGAATACCTACAACTATCTGA
TCTTTGACAAACCTGAGAAAAACAAGCAATGGGGAAAGGATTCCCTATTTAATA
AATGGTGCTGGCAAAACTGACTAGCCATATGTAGAAAGCTGAAACTGGATCCCT
TCCTTACACCTTATACAAAAATCAATTCAAGATGGATTAAAGATTTAAACGTTA
GACCTAAAACCATAAAAACCCTAGAAGAAAACCTAGGCATTACCATTCAGGACA
TAGGCATGGGCAAGGACTTCATGTCCAAAACACCAAAAGCAATGGCAACAAAAG
ACAAAATTGACAAATGGGATCTAATTAAACTAAAGAGCTTCTGCACAGCAAAAG
AAACTACCA**TCAGAGTGAACAGGCAACCTACAAAATGGGAGAAAATTTTCGCAA**
CCTACTCATCTGACAAAGGGCTAATATCCAGAATCTACAATGAACTCAAACAAA GAM25
TTTACAAAAAAAAAAAAAAA

FIG. 13B

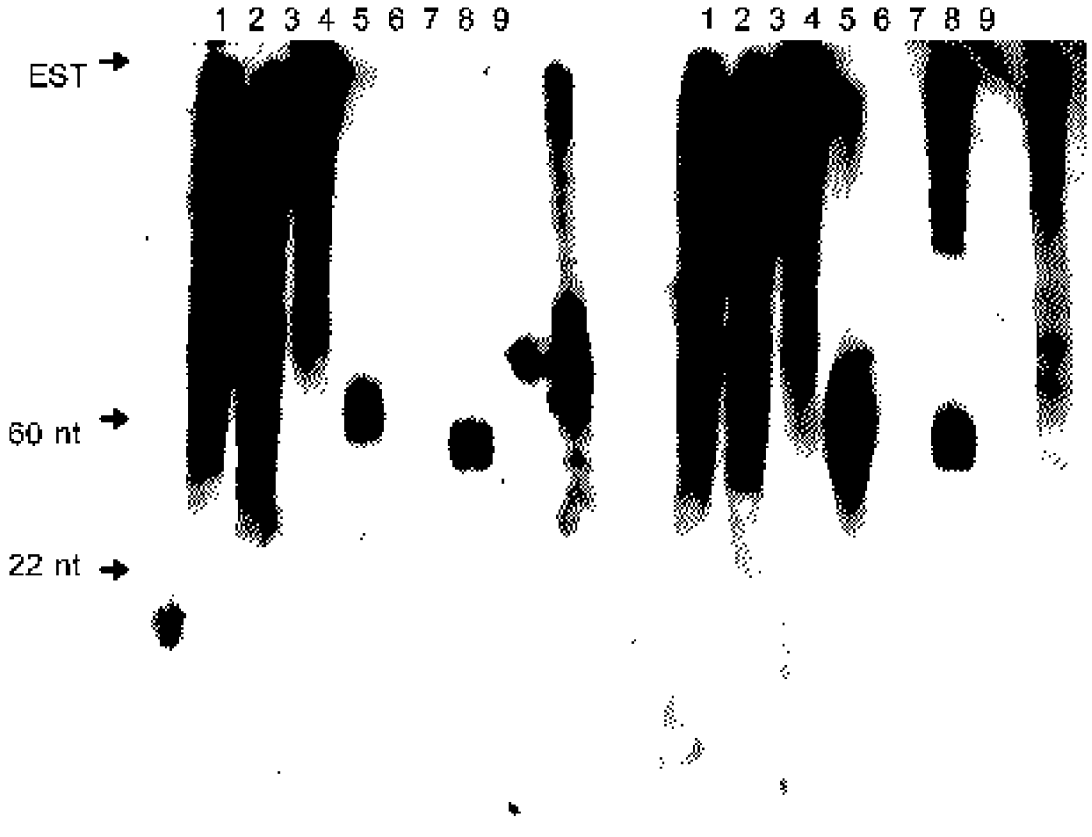
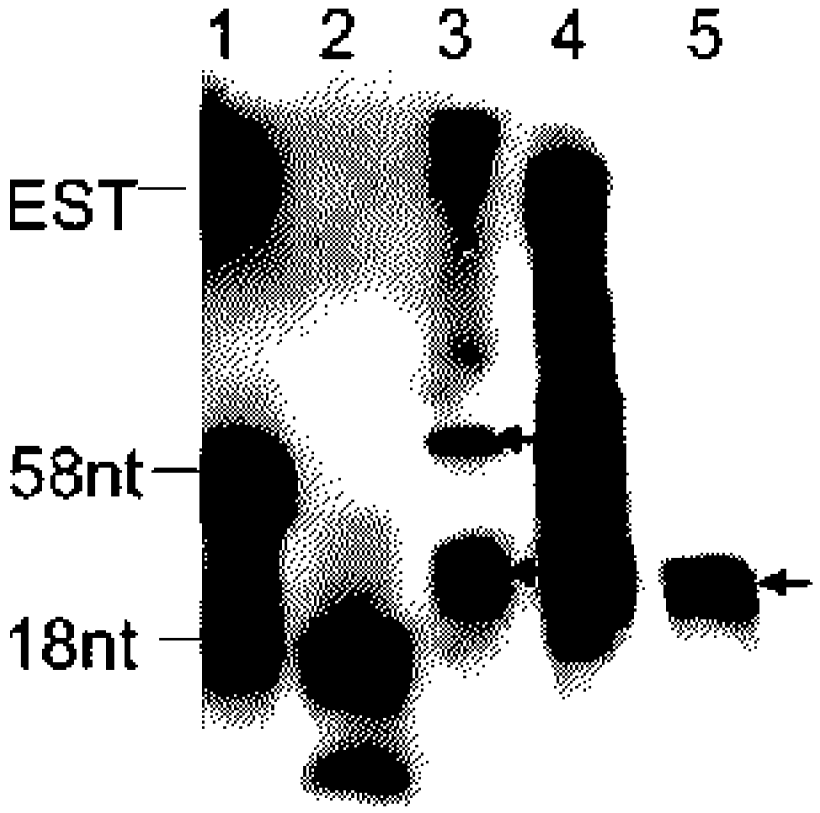


FIG. 13C



GAM25

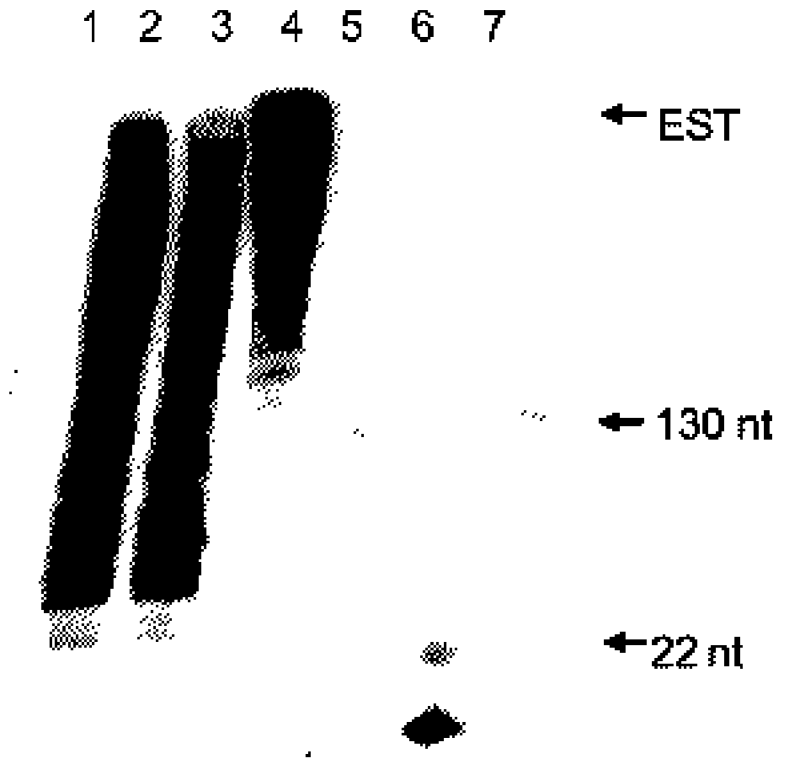
FIG. 14A

dbEST Id.1388749 (Image1020185) Sequence:

ACTCCTATCAACAGTGTAAAAGCATTCCTGTTTCTCCATAATCTTGCCAGCATCTT
TTCATTTTTTTGAATTATAGCCATTCTGACTGTTGTGAGATGGTCTCTCATTGTGG
TTTTGATTTGCATTTCTCAGATGATCAGTGATGTTGAAGTTTTTTTTGTTTGTTGGC
TGCATGTATGCCTTCTTTTGAAAAGTGTCTGTTTGTGTCCTTTGACCACTTTCTAA
TGGGGTTGAGTTTTTTTTTCTTGTAAATTTGTTTAAGTTCCTTGTAGATGCTGGAT
ATTAGACCTTTGTCAGATGGATAGAGTGCAAAAATTTTCTCCCATTCTGTAGGTTG
TCGGTTTACTCTGTTGATAGGTTCTTAATGCTGTGCAGAAGCTCTTTAGTTTAATT
AGATCCCATTTGTCAATTTTGGCTTTTGTTGCAATTGCTTTTGGCATCTTCGTCAT
GAAATCTTTGCCCTTGCCTGTGTCCTGAATGGCATTGCCTAGGTTTTCTTCCAGGA
TTTTTATAGTTTTGGGTTGTAGATTTAAGTCTTTAATCCATCTTGAGTTAACTTTT
GTATATGGCTTAAGGAAGGGGCCCCGTTTCAATTTGCTGCAAATGGCTAGCCAGTTC
TCCCAGCACCATTTATTAAATAGGGAATCTTTTCCCCATTGCTTCCTTTTGTCAGG
TTTGTCAAAGATCACATGGTTGTAGGTGTGTGGTCTTATTTCTGGGTTCTCTATTC
TGTTCCATTGGGCTATGGGCCCGTTCTGTACCACCACTATGCTGTTTTGGGTACCA
TAGTCTTGTAGAATGTTTGAAGCTGGGTAGCATGATGCCTCTAGCTTTGCTCTTCT
TGCTAAGAAATGTCTTGGCTATTTGGGCTCTTTTTTGGTTCCATATGAATTTTAAA
ATAGCTTTTTCTAGGTCTGTAAACAATGTGAATAGTAGTTTAATGGGCCTAGCATT
TAATTTACAGATTGCCTTGGGCAGTGTGGTCATTTTCACGATATTGATCCTTCCTG
TCTGTGAGCATATGTTT**TTCCATTTGTTTGTGTCATCTCTGATTTCTTTGAATAAT**
GGTTTATAGTTATCCTTGAAAAGGTCCTTCACTTTTCTTGTTAGCTGTATTCCTAG
ATATTATACTCTTCTTGTGGCAATTGTGAATGGGAGTTAATTCATGAGTTTTCTCT
CGGCTTGCCTGTTGTTGGTGTATAGGAATGCTAGTGACTTTTGCACATTGATTTTG
TATCCTGAGACTTTGTTGAAGTTGCTTATCAGCTAAGAAGTTTTTGAGCTGAGATG
ATGGAGTTTTTCTAGATATAGGATCATATCATCTGCAAACAAAGATAGTTTGACTTC
CTGTCTTCCTATTTGAATAGCTTTTCTTTCTTTCTCTTGCCTGATTGCCTTGGTGA
GAATTTCTAATACTGTGTTGAATAGGAGTGGTGAGCTCGTGCCAA

GAM 26

FIG. 14B



GAM26